The Report Committee for Weiwei Xu
Certifies that this is the approved version of the following report:

Software Implementation of Modeling and Estimation of Effect Size in Multiple Baseline Designs

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Susan N. Beretvas

James E. Pustejovsky

Software Implementation of Modeling and Estimation of Effect Size in Multiple Baseline Designs

by

Weiwei Xu, B.S.; M.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

December, 2013

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to Dr. James Pustejovsky who has walked me through the whole journey of this master report. Without his in-depth knowledge and patient guidance it would be impossible for me to finish this study. I also would like to thank Dr. Tasha Beretvas who has been my graduate advisor since I joined the statistics program. Her enthusiasm in research and broad knowledge have inspired me to study and research.

I would like to thank all the faculty members who have taught my courses, including Dr. Matt Hersh who introduced me to the statistical modeling and SAS programming and helped to build a foundation of the whole study.

A special thank-you note goes to our dearest graduate coordinator and administrative manager, Ms. Vicki Keller, who is always my go-to person for any program related questions and issues. Your prompt response, understanding of the systems, friendly yet professional dealing with issues have helped me overcome all the hurdles and continue my study successfully.

Finally I would like to thank my family and many friends who always give me encouragement and support.

Abstract

Software Implementation of Modeling and Estimation of Effect Size in Multiple Baseline Designs

Weiwei Xu, M.S. Stat.

The University of Texas at Austin, December, 2013

Supervisor: Susan N. Beretvas

A generalized design-comparable effect size modeling and estimation for multiple baseline designs across individuals has been proposed and evaluated by Restricted Maximum Likelihood method in a hierarchical linear model using R. This report evaluates the exact approach of the modeling and estimation by SAS. Three models (MB3, MB4 and MB5) with same fixed effects and different random effects are estimated by PROC MIXED procedure with REML method. The unadjusted size and adjusted effect size are then calculated by matrix operation package PROC IML. The estimations for the fixed effects of the three models are similar to each other and to that of R. The variance components estimated by the two software packages are fairly close for MB3 and MB4, but the results are different for MB5 which exhibits boundary conditions for variance-covariance matrix. This result suggests that the nlme library in R works differently than the PROC MIXED REML method in SAS under extreme conditions.

Table of Contents

List of Tables

List of Figures

Introduction

BACKGROUND

As a way of quantifying the difference between two groups, effect size is a common statistical evaluation of the strength of relationships between the outcome and treatment in meta-analysis. There are many different effect sizes for use with different types of outcome data and experimental design. In simple experiment design, Cohen's d or the standardized mean difference is widely used as the effect size (1). The basic formula to calculate the effect size has two parts. The numerator is the difference between the mean of the measurement experiment group and control group. The denominator is the standard deviation of the scores of the control group. The sign (plus or minus) indicates the direction of the relation while the value indicates the strength. Thus, an effect size near zero means on average, i.e., there is no difference between the control and experiment group; in other words, the treatment or intervention in the experiment has no effect. A positive effect size means that the experiment group performs better than the control group. A negative value means that the control group performs better than the experiment group. In both cases the larger the absolute value of the effect size, the more effective the treatment. Some researchers have agreed to an arbitrary criterion that an effect size less than 0.20 indicates weak treatment, greater than 0.8 as strong treatments and between 0.2 and 0.8 as moderate treatment (2). Although this simple categorization can provide some guidance, evaluation of the effectiveness of a treatment should be considered case by case (3).

While d-type effect size is straightforward for simple studies, in practice studies often involve more complicated designs, for which estimation of the effect size is more challenging. Single case designs (SCDs) are a class of research studies, which are becoming popular in clinical studies and other applied areas such as psychology,

counseling and education. Unlike true experiments where researchers randomly assigns participants to a control and treatment group, in SCDs each participant serves as both. The outcome variable will be repeated and systematically measured before and after intervention or treatment for each participant over time. Two important phases are involved including baseline phase and intervention phase. There are several important types of SCDs such as ABAB design, alternating treatment design, multiple baseline design, withdrawal design and reversal design, among which multiple baseline design is the most widely used method. Shadish and Sullivan (4) have compiled a master list of SCDs from 21 journals published in 2008 and found 809 single-case designs appearing in 113 studies in a variety of fields in psychology and education. Among those 54.3% of the cases used multiple baseline design and 25.1% used combination of multiple baseline design and other designs. In multiple baseline design the treatment or intervention is applied to each participant at a different time and the outcome variable is measured before and after treatment continuously until stabilized.

The traditional method to evaluate the effect of the treatment of SCDs is to draw the line graph of the outcome over time points and visually inspect the change or slope of the outcome (5). Kazdin (6) proposed that four primary criterion should be focused when inspecting the data visually. They are change in mean, change in level, change in trend and latency to change of the outcome over the transition of conditions. Researches are also looking for inferential statistical methods to detect and evaluate the treatment effect systematically and quantitatively.

The challenge of using statistical analysis is how to define and estimate the effect of the treatment in SCDs. The experimental design of the treatment or the measurement instrument of outcome could be different across studies. Though the conclusion of the treatment effect should not vary because of the study designs. Thus a design-comparable

effect size is desired to measure the effect of treatment on same metric across studies. Recently Pustejovsky (7) has successfully constructed design-comparable effect size for multiple baseline design using Hierarchical Linear Modeling (HLM). The multiple baseline design involves repeated measurements of the outcomes over time and controlled introduction of the treatment at different time for different individuals. The variance components include within-individual variation across treatment occasions and variation across individuals. The key point of the design is that the model is general enough to incorporate single case design and between-subjects randomization. Then an effect size is constructed comparably to Cohen's d-type effect size. Unlike previous studies with strong assumption that the baseline before treatment is stable and the outcome after the treatment is stable (8), Pustejovsky's work has more generalized condition by allowing trends in both baseline and after treatment phase, which may result in broader application.

This report will first introduce the construction of Pustejovsky's effect size by hierarchical linear model (HLM) and estimation the parameter by restricted maximum likelihood estimation (REML) method in R. Three variations of the model will be discussed and effect size will be estimated by REML method in SAS. The results by two different software packages will be compared and analyzed.

STATISTICAL MODEL

Pustejovsky has constructed a two-level model where level one describes a causally interpretable regression model for the i^{th} individual and level two describes the variation of the regression coefficients across cases. Causally interpretable means that the model describes not only the observed outcome data but also potential outcomes under

variations in how treatment is assigned. A general function of level one is written as following:

$$Y_{ij}(T_i) = \beta_{0i} + \beta_{1i}\, 1(j > T_i) + \beta_{2i}(j - C) + \beta_{3i}((j - T_i) \times 1(j > T_i)) + \varepsilon_{ij} \quad (1)$$

Here outcome Y_{ij} is a function of treatment assignment time T_i for the i^{th} individual at measurement time j. There are totally m individuals and N measurement occasions. For case i, T_i measurements are made during the baseline phase where $0 \leq T_i \leq N$. Then the case receives treatment and $N - T_i$ more measurements are made after the treatment. $1(j > T_i)$ is an indicator variable that equals to 0 before the treatment and 1 after receiving the treatment. We assume that the measurement occasions are equally spaced and centered at the constant C. For each individual i, the treatment may be introduced at different time, the baseline phase may have different time points, and after treatment phase may also have different time points.

The interpretation of the coefficients is described as follows.

β_{0i} : Average level of outcome at time j = C without treatment for case i

β_{1i} : Immediate change of the outcome due to the introduction of treatment for case i

β_{2i} : Linear change of the outcome per measurement occasion for case i

β_{3i} : Additional change of the outcome per measurement occasion with the treatment for case i

Simply regarding the error terms ε_{ij} as independent is not plausible since the measurements of each case are taken over time (8). Here an AR(1) model is used, where the errors are assumed to be auto-correlated and to follow a stationary, first-order auto-regressive process. The errors have expectation zero, variance σ^2, and first order autocorrelation φ, where $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) = \varphi^{|k-j|}\sigma^2$ within cases. All errors are assumed to be independent across cases where $\mathrm{Cov}(\varepsilon_{hj}, \varepsilon_{ik}) = 0$ if $h \neq i$.

The d-type design-comparable effect size is given by:

$$\delta_{AB} = \frac{E(Y_{iB}(A)) - E(Y_{iB}(N))}{\sqrt{Var(Y_{iB}(N))}} \qquad (2)$$

where effect size equals the difference between the average outcome if treatment is introduced at time A, $Y_{iB}(A)$, and if treatment is never introduced, $Y_{iB}(N)$, divided by the standard deviation of the outcome without treatment where all outcomes are measured at a fixed time B. Pustejovsky has specified five models with different group-level assumptions about the variation across cases in his study. Each model includes one varying component or a combination of several components including the intercept, treatment, time trend, and interaction between time trend and treatment. By allowing more variations in the model, more practical situations can be explained. Pustejovsky has estimated the effect size by statistical software package R for these models. In this report, a similar estimation is performed using SAS for 3 of the 5 models (MB3, MB4, and MB5), summarized as follows.

Model MB3: Varying intercepts, fixed treatment effect, fixed trends:

$$\beta_{0i} = \gamma_{00} + \eta_{0i},\ \beta_{1i} = \gamma_{10},\ \beta_{2i} = \gamma_{20},\ \beta_{3i} = \gamma_{30}, \qquad \eta_{0i} \sim N(0, \tau_0^2) \qquad (3)$$

Here β_{0i} is the intercept, β_{1i} is the treatment effect, γ_{00} can be interpreted as the average level of the outcome across individuals in the absence of the treatment, and γ_{10} is the immediate change in the outcome after intervention of treatment, γ_{20} is the change in the outcome per measurement unit in the absence of treatment, γ_{30} and η_{0i} are the additional change in the outcome per measurement unit when treatment is introduced. The slope β_{2i} is for the time trend and β_{3i} is for the time by treatment interaction. All of γ_{10}, γ_{20} and γ_{30} are constant across individuals.

Applying conditions in (3) to equation (1) we find:

$E(Y_{iB}(N)) = \gamma_{00} + \gamma_{20}$ (B-C),

$E(Y_{iB}(A)) = \gamma_{00} + \gamma_{10}+\gamma_{20}$ (B-C) + γ_{30} (B-A),

Var ($Y_{1B}(N)$ = Var ($\eta_{0i} + \varepsilon_{jB}$) = ($\tau_0^2 + \sigma^2$),

and effect size

$$\delta_{AB} = \frac{\gamma_{10}+ \gamma_{30}\ (B-A)}{\sqrt{\tau_0^2+\sigma^2}} \qquad (4)$$

Here difference (B-A) is the length of time held constant between treatment introduction and measurement, which does not depend on the choice of B because the variance is constant across measurement occasions, regardless of the pattern of treatment assignments.

Model MB4: Varying intercepts, fixed treatment effect, varying trends:

$\beta_{0i} = \gamma_{00} + \eta_{0i}$, $\beta_{1i} = \gamma_{10,}$ $\beta_{2i} = \gamma_{20} + \eta_{2i}$, $\beta_{3i} = \gamma_{30}$, (5)

Here γ_{00} is the average level of outcome across individuals without treatment and η_{0i} is the variance of the outcome for case i , γ_{10} and γ_{30} are the same as MB3, and γ_{20} is not constant as in MB3. In addition, (η_{0i}, η_{2i}) is multi-variate normally distributed, with mean (0, 0) and covariance matrix

$$T=\begin{bmatrix} \tau_0^2 & \tau_{20} \\ \tau_{20} & \tau_2^2 \end{bmatrix} \qquad (6)$$

Applying conditions in (5) to equation (1) we find:

$E(Y_{iB}(N)) = \gamma_{00} + \gamma_{20}$ (B-C),

$E(Y_{iB}(A)) = \gamma_{00} + \gamma_{10}+\gamma_{20}$ (B-C) + γ_{30} (B-A),

Var ($Y_{1B}(N)$ = Var ($\eta_{0i} + \eta_{2i}$ (B-C) + $\dot{o}_{iB}$) = τ_0^2 + (B-C)$^2\tau_2^2$ + 2(B-C) τ_{20} + σ^2,

and

$$\delta_{AB} = \frac{\gamma_{10} + \gamma_{30}\,(B-A)}{\sqrt{\tau_0^2 + (B-C)^2\tau_2^2 + 2(\mathrm{B-C})\tau_{20} + \sigma^2}} \qquad (7)$$

The effect size in (7) depends on the choice of both A and B. A simpler expression can be derived by centering at time C = B so that Var $Y_{1B}(N)$ is reduced to just (τ_0^2 + σ^2) and effect size becomes the same as in MB3. The difference (B-A) is the length of time held constant between treatment introduction and measurement. Unlike MB3, the difference (B-A) depends on the choice of B through the choice of centering point C .

Model MB5: Varying intercepts, varying trends, varying treatment-by-time interaction:

$\beta_{0i} = \gamma_{00} + \eta_{0i}$, $\beta_{1i} = \gamma_{10,}$ $\beta_{2i} = \gamma_{20} + \eta_{2i}$, $\beta_{3i} = \gamma_{30} + \eta_{3i}$, (8)

where (η_{0i}, η_{2i}, η_{3i}) is multi-variate normally distributed, with mean (0, 0, 0) and covariance matrix

$$\mathrm{T}=\begin{bmatrix} \tau_0^2 & \tau_{20} & \tau_{30} \\ \tau_{20} & \tau_2^2 & \tau_{32} \\ \tau_{30} & \tau_{32} & \tau_3^2 \end{bmatrix} \qquad (9)$$

In this model, in the absence of treatment, cases vary in their average levels of the outcome and slope of change over time. The treatment also has variable effects by influencing the interaction term. Applying conditions in (8) to equation (1) we find the effect size is the same as equation (7).

EFFECT SIZE ESTIMATION

In the above three HLM models, we can denote the (p x 1) vectors of fixed effects by $\gamma = (\gamma_{00}, \ldots, \gamma_{(p-1)0})^T$ and the (r x 1) vectors of the variance components by $\omega = (\tau_0^2, \ldots, \varphi, \sigma^2)^T$. Effect size can then be simply expressed as a matrix operation:

$$\delta_{AB} = \frac{P^T\gamma}{\sqrt{R^T\omega}} \qquad (10)$$

where p is the number of fixed effects and r is the number of all variance components including auto-correlation coefficient φ, within-case variance σ^2 and random effect covariance τ. For example, in MB3, take $\gamma = (\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30})^T$, P = (0, 1, 0, (B-A))^T, take $\omega = (\tau_0^2, \varphi, \sigma^2)^T$, R = (1, 0,1)^T, the effect size calculated by equation (10) is the same as in equation (4).

The RML method then can be applied by estimating $\hat{\gamma}$, $\hat{\omega}$ and $\hat{\delta}_{AB}$. Based on the theorem given in Hedges (7), the distribution of $\hat{\delta}_{AB}$ can be approximated by a constant *k* times a non-central t distribution with *v* degrees of freedom:

$$k = \sqrt{\frac{P^T C(\hat{\gamma}) P}{r^T \hat{\omega}}} \tag{11}$$

$$v = \frac{2(r^T \hat{\omega})^2}{r^T C(\hat{\omega}) r} \tag{12}$$

An adjusted effect size is given by

$$g_{AB} = j(v) \times \hat{\delta}_{AB} \tag{13}$$

where

$$j(v) = 1 - 3/(4v - 1) \tag{14}$$

The variance of g_{AB} is approximately

$$Var(g_{AB}) \approx j(v)^2 [\frac{vk^2}{v-2} + \hat{\delta}_{AB}^2 \left(\frac{v}{v-2} - \frac{1}{j(v)^2}\right)] \tag{15}$$

Pustejovsky has implemented the RML estimation using the nlme package in R to estimate constant *k* and degree of freedom *v* to calculate the modified effect size. This report will use the Proc Mixed with REML method in SAS 9.2 to implement the

hierarchical linear models (MB3, MB4 and MB5) and calculate the effect size by matrix operation package Proc IML.

Implementation of the Methods by SAS

SAS PROC MIXED

Mixed procedure in SAS can be used to fit a variety of data and make statistical inference. Unlike the Generalized Linear Model (GLM) procedure for the standard linear model, the data in Mixed model are permitted to have correlations and non-constant variability; thus variance and covariance of the data can be modeled. There are three primary assumptions underlying the analysis by Proc Mixed. The first one is that the data are normally distributed (Gaussian). The second is that the means of the outcome are linear to a set of parameters noted as fixed effects. The third assumption is that the variance and covariance of the data are in terms of a different set of parameters noted as covariance parameters. The fixed effects parameters are associated with known explanatory variables which are the same as in traditional linear models. The unknown random effects provide additional variability to the data. The variances of the random effects parameters become the covariance parameter of the structure. The mixed procedures are widely used in the following two typical scenarios:

First one is called nested study. The experimental units can be grouped into clusters and the data from a common cluster are correlated. For example, if students are the experiment unit, they can be grouped into classes, which in turn can be grouped into schools and then districts. Each level of this hierarchy can introduce an additional variability and correlation in the HLM model.

The second one is called longitudinal study. Repeated measurements are taken on the same experimental unit over time. The measurements are correlated or exhibit variability. This model is widely used in pharmaceutical clinical study, psychology and social science. Single case design study or multiple baseline study also conducts repeated

measurements over time for each individual which can be viewed as a special case of longitudinal study. Therefore the mixed procedure is used in this study.

Proc Mixed fits the structure using method of restricted maximum likelihood (REML). It constructs an objective function associated with REML and maximizes it over all unknown parameters using ridge-stabilized Newton-Raphson (NR) algorithm (9). The advantage of NR algorithm over Expectation-Maximum (EM) has been presented in a previous study (10). The complicated mathematics of the estimation procedure is not discussed in this section.

SUMMARY OF SCHUTTE DATA

Schutta, Malouff, and Brown (11) conducted a study to evaluate adult's prolonged fatigue problem by emotion-focused therapy using a multiple baseline across individuals. Before the treatment, cases were measured weekly, some for 2 weeks, some for 5 weeks and some for 8 weeks. Then treatments were applied. Cases were measured weekly again. The total length of the measurement is different for cases and it ranged from 1 to 7 weeks. The longest case was 8 weeks of baseline and 7 weeks after treatment with total 15 time points. The shortest case was 5 weeks of baseline and 1 week after treatment with total 6 time points. The fatigue measurement was a self-reported scale from 1 to 63. A total of 12 cases and 136 time points will be evaluated in this report.

APPLICATIONS

Previously Pustejovsky has used nlme package in R to estimate the effect size of different models. This report will replicate the calculation with Proc MIXED /REML method and Proc IML package in SAS. I choose the three models (MB3, MB4 and MB5) described above to study the Schutte data and to estimate the effect size. The models differ only in whether the case-level regression specification are assumed to be constant

or allowed to vary. The model has three fixed factors including time trend, treatment and the interaction between time trend and treatment. The effect size parameter depends on the choice of time-points A and B for describing the hypothetical between-subjects design. For all these three models I use A=2, B=9 and C=9. By A=2, the treatment would be applied after the second measurement occasion in a hypothetical experiment. By B=9, the effect size measures the effect of B-A=7 weeks of the treatment, which is the maximum length recorded in the data. By C=9, I simplify the calculation by centering the weekly trend at 9 weeks. Therefore the case-level interpretation is corresponding to the average level of the outcome after 9 weeks without treatment.

Next I present the SAS codes for all three models and discuss the results in comparison with the results from R, as calculated by Pustejovsky.

MB3 Code and Results

Model 3 has varying intercept, fixed treatment effect, fixed trends and fixed interaction as in equation (3).

$\beta_{0i} = \gamma_{00} + \eta_{0i}$, $\beta_{1i} = \gamma_{10}$, $\beta_{2i} = \gamma_{20}$, $\beta_{3i} = \gamma_{30}$, $\eta_{0i} \sim N(0, \tau_0^2)$ (3) The SAS code is illustrated and explained in three parts. Figure 1 is for data input, Figure 2 is modeling by PROC MIXED and output data reference, and Figure 3 is matrix operation by PROC IML to estimate the unadjusted and adjusted effect size .

```
OPTIONS formdlim='_' ls=100;

FILENAME io 'U:\My Documents\My SAS Files\9.2\Effect_Size';

DATA schutte;
        INFILE io(Schutte.csv) dlm='2C0D'x dsd missover lrecl=10000
        firstobs=2;

        INPUT Num case Week Fatigue Trt CaseID Treatment$ Constant
        Trt_Week;

PROC PRINT DATA=schutte (obs=15); RUN;
```

Figure 1: SAS code for MB3, part 1: Data input.

The first part of the code is to get the raw data Schutte.csv into SAS and create SAS data set schutte with multiple variables specified in the INPUT statement. The raw data is saved in the directory of 'U:\My Documents\My SAS Files\9.2\Effect_Size'. The first 15 observations are checked to see if the data are read correctly. After that the modeling part is illustrated in part 2.

```
ODS TRACE ON;
        ODS OUTPUT
                CovParms=w1 (keep=covparm estimate)
                asycov=w2 (drop=row)
                SolutionF=w3 (keep= effect estimate)
                covb=w4 (drop =row);

        PROC mixed DATA=schutte METHOD=REML noclprint asycov
        covtest;
        CLASS case;
        MODEL Fatigue = Trt Week Trt_Week /SOLUTION covb;
        REPEATED /SUB=case TYPE = ar(1);
        RANDOM intercept  / SUB= case TYPE=un;
RUN;
ODS TRACE OFF;
```

Figure 2: SAS code for MB3, part 2: Proc Mixed modeling

Part 2 of the code is the main body of the modeling with estimation of all the fixed and random effects. The pair of ODS TRACE ON and OFF statement help to track the table names from the output in the log file of SAS. ODS output will allow the direct reference and modification of the output tables and assign to new dataset for future calculation. The KEEP and DROP option can specify the selection of certain columns in the data set. In the PROC MIXED statement we tell SAS to use data set Schutte created in part 1 as input, use method REML as estimation method, print out standard deviation of estimated parameter by option covtest, and display the covariance matrix of parameter estimates by option asycov. The CLASS statement tells SAS to treat case as a categorical or classification variable. The MODEL statement together with options is the core part of the code. It specifies that Fatigue is the outcome variable, three independent variables Trt, Week, and the interaction term Trt_Week are the three fixed effects. The option SOLUTION displays the parameter estimate in the output solutionF table according to the ODS TRACE line. The option covb displays the covariance matrix of the fixed effect parameter estimate in the output covb table. REPEATED statement estimates the variance-covariance matrix of the random residuals ε_{ij}, which can be viewed as the within-subject variance. The TYPE option tells the structure is auto-regressive AR(1). The RANDOM statement defines the random effects. In MB3 only intercept ($\beta_{0i} = \gamma_{00} + \eta_{0i}$) is varying. The output covParms table shows the estimate. The TYPE option tells the variance-covariance matrix of the random effects is unstructured and output asycov table shows the covariance matrix of the estimates. Now SAS data sets w1,w2,w3 and w4 are

generated by PROC MIXED procedure according to the ODS TRACE and OUTPUT. They will be used in the code part 3 as matrices for further calculation.

```
PROC IML;
        P={0,1,0,7};
        USE w3; READ all INTO Gamma;
        USE w4; READ all INTO C_Gamma;

        R={1,0,1};
        USE w1; READ all INTO Omiga;
        USE w2; READ all INTO C_Omiga;

        Trt7= T(P)*Gamma; Std_Trt= sqrt(T(P)*C_Gamma*P);

        Tot_Var= T(R)* Omiga;
        Std_Var=sqrt(T(R)*C_Omiga*R);

        Delta_AB=T(P)*Gamma/sqrt(T(R)* Omiga);

        k=sqrt((T(P)*C_Gamma*P)/(T(R)* Omiga));
        v=2*((T(R)*Omiga)**2)/(T(R)*C_Omiga*R);
        j_v=1-3/(4*v-1);
        G_AB=j_v*Delta_AB;
        V_G_AB=j_v**2*(v*(k**2)/(v-2)+(Delta_AB**2)*(v/(v-2)-1/(j_v**2)));
        Std_Vgab=sqrt(V_G_AB);

        PRINT p,Gamma,C_Gamma,R,Omiga,C_Omiga;
        PRINT Tot_Var, Std_Var,Trt7,Std_Trt,Delta_AB,k,v, j_v,
G_AB,V_G_AB,Std_Vgab;

QUIT;
RUN;
```

Figure 3: SAS code for MB3, part 3: Matrix operation by Proc IML.

The part 3 of the code is to use the corresponding SAS output data sets to set up the fixed effects and random effect and their covariance matrices and calculate the effect size. PROC IML package in SAS can perform general matrix operation such as

multiplication, division, square root, etc. P and R vector can be named directly. The combination of USE and READ commands can set up matrices using the data sets produced in the PROC MIXED part. The constant *k* and degree of freedom *v*, effect size $\hat{\delta}_{AB}$, adjusted effect size g_{AB} and variance of g_{AB} are then calculated by PROC IML package according to the definitions from equation (10) to equation (15).

Table 1. Model MB3 estimates for Schutte data by R and SAS

Parameter	**R(J.P) Estimate**	**(s.e.)**	**SAS (W.X) Estimate**	**(s.e.)**
Variance components				
Autocorrelation ($\hat{\phi}$)	0.81	(0.02)	0.81	(0.09)
Within-case var. ($\hat{\sigma}^2$)	99.00	(6.20)	99.01	(44.74)
Between-case var. ($\hat{\tau}_0^2$)	14.77	(27.16)	14.75	(39.51)
Total variance ($\hat{\tau}_0^2 + \hat{\sigma}^2$)	113.77	(27.14)	113.76	(29.99)
Fixed effects				
Intercept ($\hat{\gamma}_{00}$)	52.93	(4.42)	52.93	(4.42)
Treatment ($\hat{\gamma}_{10}$)	-1.37	(1.97)	-1.37	(1.97)
Weekly trend ($\hat{\gamma}_{20}$)	0.49	(0.62)	0.49	(0.62)
Trt. × Trend ($\hat{\gamma}_{30}$)	-1.90	(0.94)	-1.90	(0.94)
Trt. effect after 7 weeks ($\mathbf{p}^T\hat{\boldsymbol{\gamma}}$)	-14.65	(6.34)	-14.65	(6.34)
Effect size				
Unadjusted ($\hat{\delta}_{AB}$)	-1.37		-1.37	
Adjusted (g_{AB})	-1.34	(0.63)	-1.34	(0.63)
Degrees of freedom (*v*)	35.15		28.77	
Constant κ	0.59		0.59	
Log-likelihood	-435.1		-435.1	
Akaike Info. Criterion	884.2		876.2	

The full SAS code is given in the APPENDIX section. In summary, the various datasets from the output of PROC MIXED in SAS can be converted to matrices easily and the

matrix operations in the package are straightforward. The sample calculation with all matrix elements is demonstrated in MB4 section since MB4 has more complicated variance-covariance matrix for random effects. Table 1 shows the estimated results by MB3 from R and SAS side by side.

The results from SAS and R are very close. The fixed effects, unadjusted or adjusted effect size and Log-likelihood are nearly identical. The coefficients for variance components are the same, but the standard deviations of the coefficients have variations which may result from the different algorithms for estimating the covariance matrix of variance components in SAS and R. Previous investigation has showed that MB3 is a poor fit to the data.

MB4 Code and Results

Model 4 has varying intercept, fixed treatment effect, varying trends and fixed interaction as in equation (5):

$$\beta_{0i} = \gamma_{00} + \eta_{0i},\ \beta_{1i} = \gamma_{10},\ \beta_{2i} = \gamma_{20} + \eta_{2i},\ \beta_{3i} = \gamma_{30}$$

```
PROC MIXED data=schutte method=REML noclprint asycov covtest;
CLASS case;
MODEL Fatigue = Trt Week Trt_Week /solution covb;
REPEATED /sub=case type = ar(1);
RANDOM intercept Week / sub= case type=un;
```

Figure 4: SAS code for MB4, PROC MIXED part

The code for MB4 in Figure 4 is almost identical as MB3 except for the RANDOM statement. In MB4 the baseline time trend is allowed to vary randomly across

cases besides the variation of the intercept. In the SAS code I add Week to the RANDOM command line.

The RANDOM line in MB3 is:

```
RANDOM intercept / sub= case type=un;
```

And the RANDOM line in MB4 is:

```
RANDOM intercept Week / sub= case type=un;
```

The full SAS code with explanation is attached in the Appendix section.

The results from R and SAS are again very close for MB4, as listed in Table 2. Previous study has suggested MB4 has improved fit than MB3.

Some examples of matrix operation are listed here. For the given choice of A, B and centering point C, the fixed effects vector P is

$$\mathrm{P} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 7 \end{pmatrix},$$

Fixed effects γ is from the Estimate of the SolutionF table from SAS output. SolutionF table describes the fixed effects solution vector.

$$\gamma = \begin{pmatrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{pmatrix} = \begin{pmatrix} 50.29 \\ -0.54 \\ 0.20 \\ -1.63 \end{pmatrix},$$

The covariance matrix of the fixed effects $C(\gamma)$ is from the CovB table from SAS output. CovB table describes the covariance matrix of fixed-effects parameter estimates.

$$C(\gamma) = \begin{bmatrix} 16.594454 & -1.841364 & 2.2172766 & -1.513235 \\ -1.841364 & 3.0683475 & -0.215684 & -0.204787 \\ 2.2172766 & -0.215684 & 0.3796935 & -0.240505 \\ -1.513235 & -0.204787 & -0.240505 & 0.4293614 \end{bmatrix}$$

The variance vector r is

$$\mathrm{r} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

random effects ω is from the estimate of the CovParm table from SAS output.

CovParm table describes estimated covariance parameters.

$$\omega = \begin{pmatrix} \tau_0^2 \\ \tau_{20} \\ \tau_2^2 \\ \varphi \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} 95.712724 \\ 11.209038 \\ 1.9938047 \\ 0.3985966 \\ 29.394447 \end{pmatrix}$$

The covariance matrix of the random effects C(ω)is from the Asycov table from SAS output. Asycov table describes asymptotic covariance matrix of covariance parameters.

$$C(\omega) = \begin{bmatrix} 2161.4019 & 260.53245 & 32.063919 & -0.643367 & -32.43897 \\ 260.53245 & 40.545795 & 6.0490924 & -0.10067 & -4.93707 \\ 32.063919 & 6.0490924 & 1.1653079 & -0.02753 & -1.366049 \\ -0.643367 & -0.10067 & -0.02753 & 0.0160823 & 0.6626841 \\ -32.43897 & -4.93707 & -1.366049 & 0.6626841 & 42.932269 \end{bmatrix}$$

Trt. Effect after 7 weeks = P^T x γ = -11.96606 ;

Standard Deviation(s.e) of the Trt. effect after 7 weeks

$=\sqrt{\mathrm{p^T x\ C(\gamma) x P}} = 4.6086912$;

Total variance = r^Tx ω = 125.10717

s.e.of the Total variance = $\sqrt{\mathrm{r^T x\ C(\omega) x\ r}}$ = 46.254256

Effect size $\delta_{AB} = \frac{P^T\gamma}{\sqrt{r^T\omega}}$ = -1.069819

Constant $k = \sqrt{\frac{P^T C(\hat{\gamma})P}{r^T\hat{\omega}}}$ = 0.4120373

Degree of freedom $v = \frac{2(r^T\hat{\omega})^2}{r^T C(\hat{\omega})r}$ = 14.631572

$j(v) = 1 - 3/(4v - 1)$ = 0.9478499

Adjusted effect size $g_{AB} = j(v) \times \hat{\delta}_{AB}$ = -1.014027

The variance of g_{AB} is approximately

$Var(g_{AB}) \approx j(v)^2[\frac{vk^2}{v-2} + \hat{\delta}_{AB}^2\left(\frac{v}{v-2} - \frac{1}{j(v)^2}\right)]$ = 0.2232259

And the s.e of the adjusted effect size $g_{AB} = 0.4724678$

Some estimates are interpreted as follows. First, $\hat{\gamma}_{10}$ = -0.54, meaning that fatigue is immediately lowered by 0.54 point after application of the treatment. With $\hat{\gamma}_{30}$ = -1.63, fatigue will be lowered by additional 1.63 points per week after the treatment. Together, fatigue will be lowered by 11.97 points seven weeks after treatment. Since $\hat{\gamma}_{20}$ = 0.20, the average baseline trend is 0.20, which is close to zero. The variation of the baseline trend across cases ($\hat{\tau}_2^2$) is 1.99. Total variance is 125.11 and the majority is coming from the between-case variance which is 95.71. Within-case variation is 29.39, assumed to be constant for all cases. The unadjusted effect size is -1.07 of standard deviation unit and adjusted effect size is -1.01 of standard deviation unit with approximate variance 0.47. The average treatment effect is to lower the fatigue by 1 unit of standard deviation. Since the effect size is design-comparable it can be compared to that of other experiment designs or models.

Table 2. Model MB4 estimates for Schutte data by R and SAS

Parameter	R (J.P) Estimate	(s.e.)	SAS (W.X) Estimate	(s.e.)
Variance components				
Autocorrelation ($\hat{\phi}$)	0.40	(0.09)	0.40	(0.13)
Within-case var. ($\hat{\sigma}^2$)	29.39	(4.16)	29.39	(6.55)
Between-case var. ($\hat{\tau}_0^2$)	95.71	(46.79)	95.71	(46.49)
Case-trend covariance ($\hat{\tau}_{20}$)	11.21	(6.36)	11.21	(6.37)
Trend variance ($\hat{\tau}_2^2$)	1.99	(1.07)	1.99	(1.08)
Total variance ($\hat{\tau}_0^2 + \hat{\sigma}^2$)	125.11	(46.76)	125.11	(46.25)
Fixed effects				
Intercept ($\hat{\gamma}_{00}$)	50.29	(4.07)	50.29	(4.07)
Treatment ($\hat{\gamma}_{10}$)	-0.54	(1.75)	-0.54	(1.75)
Weekly trend ($\hat{\gamma}_{20}$)	0.20	(0.62)	0.20	(0.62)
Trt. × Trend ($\hat{\gamma}_{30}$)	-1.63	(0.66)	-1.63	(0.65)
Trt. effect after 7 weeks ($\mathbf{p}^T\hat{\boldsymbol{\gamma}}$)	-11.97	(4.61)	-11.97	(4.61)
Effect size				
Unadjusted ($\hat{\delta}_{AB}$)	-1.07		-1.07	
Adjusted (g_{AB})	-1.01	(0.47)	-1.01	(0.47)
Degrees of freedom (ν)	14.32		14.63	
Constant κ	0.41		0.41	
Log-likelihood	-429.0		-429	
Akaike Info. Criterion	876.0		868.0	

MB5 Code and Results

Model 5 has varying intercept, fixed treatment effect, varying trends and varying interaction as in equation (8):

$\beta_{0i} = \gamma_{00} + \eta_{0i}$, $\beta_{1i} = \gamma_{10,}$ $\beta_{2i} = \gamma_{20} + \eta_{2i}$, $\beta_{3i} = \gamma_{30} + \eta_{3i}$,

By following the codes in MB3 and MB4, we can just change the RANDOM statement by adding a Trt_Week interaction term to the random effect as following:

```
RANDOM intercept Week Trt_Week/ sub= case type=un;
```

Instead of output similar results as in MB3 and MB4, an error message is received, stating that the iteration does not converge. In this model we have three random effects which means there are six terms in the covariance matrix. With only 12 cases in the dataset the REML method will be limited. The convergence problem is solved by changing the covariance matrix type from UN to UNR. The new random statement is:

```
RANDOM intercept Week Trt_Week/ sub= case type=unr;
```

The MIXED documentation in SAS (12) describes the difference between TYPE=UN and TYPE=UNR covariance structures. When TYPE=UN, the variances are nonnegative, and the covariances are unstructured. On the other hand when TYPE=UNR, the structure fits the same model as the TYPE=UN option but with a different parameterization. The non-diagonal elements are correlations between the corresponding measurements instead of covariance and the absolute value of the correlation is constrained to be not greater than 1. We can change the TYPE option to UN for both

MB3 and MB4 to achieve same estimation results if the nonnegative, unconstrained covariance structure is not sought. The results for MB5 with covariance type=UNR are listed in Table 3.

Table 3. Model MB5 estimates for Schutte data by R and SAS

Parameter	R (J.P) Estimate	(s.e.)	SAS(W.X) Estimate	(s.e.)
Variance components				
Autocorrelation ($\hat{\phi}$)	0.24	(0.11)	0.24	(0.12)
Within-case var. ($\hat{\sigma}^2$)	22.54	(3.35)	22.64	(3.93)
Between-case var. ($\hat{\tau}_0^2$)	38.30	(33.00)	75.71	(76.77)
Case-trend covariance ($\hat{\tau}_{20}$)	0.38	(3.40)	0.93	(0.97)
Trend variance ($\hat{\tau}_2^2$)	0.15	(0.54)	0.16	(0.27)
Case-Trt. × Trend cov. ($\hat{\tau}_{30}$)	1.74	(7.42)	-0.17	(0.55)
Trend-Trt. × Trend cov. ($\hat{\tau}_{32}$)	0.67	(0.96)	1	(0)
Trt. × Trend variance ($\hat{\tau}_3^2$)	3.01	(2.70)	2.84	(2.11)
Total variance ($\hat{\tau}_0^2 + \hat{\sigma}^2$)	60.84	(32.74)	98.35	(77.50)
Fixed effects				
Intercept ($\hat{\gamma}_{00}$)	50.53	(2.82)	50.57	(3.39)
Treatment ($\hat{\gamma}_{10}$)	0.03	(1.60)	0.07	(1.61)
Weekly trend ($\hat{\gamma}_{20}$)	0.22	(0.36)	0.23	(0.38)
Trt. × Trend ($\hat{\gamma}_{30}$)	-1.67	(0.74)	-1.77	(0.73)
Trt. effect after 7 weeks ($\mathbf{p}^T\hat{\boldsymbol{\gamma}}$)	-11.67	(5.18)	-12.33	(5.17)
Effect size				
Unadjusted ($\hat{\delta}_{AB}$)	-1.50		-1.24	
Adjusted (g_{AB})	-1.33	(0.83)	-0.93	(1.06)
Degrees of freedom (v)	6.90		3.22	
Constant κ	0.66		0.52	
Log-likelihood	-424.5		-424.5	
Akaike Info. Criterion	873.0		863.0	

Although the estimated fixed effects by two codes are similar to each other and to MB4, the estimated variance components and effect size are quite different for MB5. The between-case variance in MB5 is 38.3 by R but 75.71 by SAS while it is 95.71 in MB4 and 99 in MB3 by both codes. Hence the total variance and standard deviation of the total variance are quite different, which also lead to larger effect size and adjusted effect size. The reduction of between-case variance from MB3 to MB4 or from MB4 to MB5 is because the introduction of the variance of trend or interaction so that the remaining unexplained variance is decreased. The difference of between-case variation between R and SAS may be a result of the different algorithms of restricted maximum likelihood estimation of variance-covariance components. For the three models with 12 cases, relaxed restriction on the boundary condition for variance-covariance matrix is necessary for the REML package in SAS to conduct estimation.

Conclusion

Effect size is very useful in studies evaluating the causal effect of intervention in social science, psychology and clinical study. Pustejovsky's method to derive and estimate design-comparable effect size for treatment effects in multiple baselines is studied in this report. A successful estimation of effect size involves model selection, effect size construction and estimation. Three models with same fixed effects but different random effects are implemented by SAS PROC MIXED function with REML estimation. The Unadjusted effect size and adjusted effect size are then calculated by SAS matrix operation package PROC IML. The output and results from SAS are comparable to those by R for fixed effects in all models. The covariance estimation sometimes show different results at the boundary condition, which may be caused by different REML algorithm.

Future studies may be conducted by evaluating the proposed effect size estimation by other software packages such as SPSS and HLM. One could monitor the capacity of convergence at extreme boundary conditions to explore different restricted maximum likelihood estimation algorithm and evaluate fitting data. One could also study more complicated data structure with more data hierarchy to generalize the proposed estimation and widen the application. One may also study different estimation methods other than REML such as Bayes' estimator or MCMC algorithm. With the introduction of estimation of the design-comparable effect size for multiple baseline designs, more designs may be explored in future studies.

APPENDIX

1. SAS code for MB4 with comment line

```
OPTIONS formdlim='_' ls=100;

FILENAME io 'U:\My Documents\My SAS Files\9.2\Effect_Size';
/*read schutte.csv file, change data name to model3, name of last column change
to Trt_Week)*/

DATA schutte;
       INFILE io(Schutte.csv) dlm='2C0D'x dsd missover lrecl=10000 firstobs=2;
       INPUT Num case Week Fatigue Trt CaseID Treatment$ Constant Trt_Week;
/*print out a sample observations to make sure the file is read correctly*/
PROC PRINT DATA=schutte (obs=15); RUN;
/*ODS trace on and off pair helps to track the table names from the output in
the log file of SAS*/
/*ODS output will allow the direct reference and modification of the output
tables and assigned to new dataset.*/
/*The keep and drop option is to modify the dataset for further usage*/
ODS TRACE ON;
       ODS OUTPUT
              CovParms=w1 (keep=covparm estimate)/*CovParms is table of
estimation of the random parameters used for Omiga*/
              asycov=w2 (drop=row) /*asycov is the table of covariance matrix
for random parameters used for C_Omiga*/
              SolutionF=w3 (keep= effect estimate)/*SolutionnF is the table of
estimation of fixed effects used for Gamma*/
              covb=w4 (drop =row); /*covb is the table of covariance matrix for
fixed effects used for C_Gamma*/
/* Proc Mixed is to specify the regression model. */
/*The fixed effects are in the Model line, The random effects are in the random
line*, the within-case variance is in repeated line*/
PROC mixed DATA=schutte METHOD=REML noclprint asycov covtest;
       CLASS case;
       MODEL Fatigue = Trt Week Trt_Week /SOLUTION covb;
       REPEATED /SUB=case TYPE = ar(1);
       RANDOM intercept Week / SUB= case TYPE=un;

RUN;
ODS TRACE OFF;

/*print out the dataset for matrix reference*/
PROC PRINT DATA=w1;
```

```
PROC PRINT DATA=w2;
PROC PRINT DATA=w3;
PROC PRINT DATA=w4;
RUN;

/* PROC IML is the package in SAS for matrix operation*/
PROC IML;
       P={0,1,0,7}; /*define P vector as of day 7 after trearment*/
       USE w3; READ all INTO Gamma; /*define matrix of fixed effects as Gamma*/
       USE w4; READ all INTO C_Gamma; /* define matrix of covariance of fixed
effects as C-Gamma*/

       R={1,0,0,0,1}; /*define r vector*/
       USE w1; READ all INTO Omiga; /*define the matrix of random effects as
Omiga*/
       USE w2; READ all INTO C_Omiga; /*define the matrix of covariance of
random effects as C_Omiga*/

       Trt7= T(P)*Gamma; Std_Trt= sqrt(T(P)*C_Gamma*P);/*estimate the treatment
effect and standard deviation after 7 weeks*/
       Tot_Var= T(R)* Omiga; /*estimate total variance*/
       Std_Var=sqrt(T(R)*C_Omiga*R); /*estimate the standard deviation of the
total variance*/
       Delta_AB=T(P)*Gamma/sqrt(T(R)* Omiga); /*estimate the unadjusted effect
size*/
       k=sqrt((T(P)*C_Gamma*P)/(T(R)* Omiga)); /*estimate constant k*/
       v=2*((T(R)*Omiga)**2)/(T(R)*C_Omiga*R); /*estimate degree of freedom v*/
       j_v=1-3/(4*v-1); /*estimate j*/
       G_AB=j_v*Delta_AB; /*estimate the adjusted effect size*/
       V_G_AB=j_v**2*(v*(k**2)/(v-2)+(Delta_AB**2)*(v/(v-2)-1/(j_v**2)));
/*estimate the variance of adjusted effect size*/
       Std_Vgab=sqrt(V_G_AB); /*estimate the standard deviation of effect size*/

       PRINT p,Gamma,C_Gamma,R,Omiga,C_Omiga;
       PRINT Tot_Var, Std_Var,Trt7,Std_Trt,Delta_AB,k,v, j_v,
G_AB,V_G_AB,Std_Vgab;

QUIT; /*Quit from PROC IML*/
RUN;
```

The only difference for MB3 and MB5 to MB4 is the RANDOM line in PROC MIXED and R matrix in PROC IML:

SAS code for MB3:

```
RANDOM intercept / SUB= case TYPE=un;
R={1,0,1};
```

SAS code for MB5:

```
RANDOM intercept Week Trt_Week/ SUB= case TYPE=unr;
R={1,0,0,0,0,0,0,1};
```

2. PROC MIXED Options Summary

The following summarize most options in PROC MIXED Statement.

DATA=	Specifies input data set
METHOD=	Specifies the estimation method
ASYCOV	Displays asymptotic covariance matrix of covariance parameter estimates
NOCLPRINT	Suppresses "Class level information" completely or in parts
COVTEST	Produces asymptotic standard errors for the covariance parameter estimates
MODEL	Name the single dependent variable and fixed effects
SOLUTION	Displays the fixed-effects parameter estimates
COVB	Displays the covariance matrix of the fixed-effects parameter estimates
REPEATED	Specify the R^1 matrix in mixed model
RANDOM	Defines the random-effects and G^2 matrix in mixed model
TYPE	Specify the covariance structure

[1] R matrix: the variance-covariance matrix for random residuals, can be viewed as within-subject variance-covariance matrix

[2] G matrix: the variance-covariance matrix for random effects, can be viewed as between-subject variance-covariance matrix

References

3. Cooper, H. M. (2009). Hypotheses and problems in research synthesis. In H. M. Cooper, L. V Hedges, & J. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed., pp. 19–35). New York, NY: Russell Sage Foundation.
4. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates
5. Cohen, J, A power primer. *Psychological Bulletin*, Vol 112(1), Jul 1992, 155-159. doi: 10.1037/0033-2909.112.1.155
6. Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43*(4), 971–80. doi:10.3758/s13428-011-0111-y
7. Mckay D. (2008), *Handbook of research methods in abnormal and clinical psychology*, chapter 22, page 337-350 (Nock M. et al)
8. Kazdin, A. E. (1982). Single-case research designs:Methods for clinical and applied settings. NewYork: Oxford University Press
9. Pustejovsky, J. E. (2013). *Operationally comparable effect sizes for meta-analysis of single-case research*. Northwestern University
10. Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. doi:10.3102/1076998606298043
11. Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer Verlag.
12. Lindstrom, M. J. and Bates, D. M. (1988). "Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measure data," *Journal of the American Statistical Association*, 83, 1014-1022
13. Schutte, N. S., Malouff, J. M., & Brown, R. F. (2008). Efficacy of an emotion-focused treatment for prolonged fatigue. *Behavior Modification*, *32*(5), 699–713. doi:10.1177/0145445508317133
14. SAS/STAT(R) 9.2 User's guide, 2nd edition, Random and Repeated statement, TYPE= option, Table 56.13 and Table 56.14. https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mixed_sect019.htm#statug.mixed.mixedspcovstruct